

シーン中の3次元点群の投影誤差による AR/MR 向けカメラキャリブレーションの精度評価

林 将之[†] 北原 格[‡] 亀田 能成[‡] 大田 友一[‡]

[†] 筑波大学 大学院システム情報工学研究科 〒305-8573 茨城県つくば市天王台 1-1-1

E-mail: [†] mhayashi@image.iit.tsukuba.ac.jp, [‡] {kitahara, kameda, ohta}@iit.tsukuba.ac.jp

あらまし 拡張現実感や複合現実感において幾何学的整合性の確保は重要な課題であり、様々なカメラキャリブレーション方式が提案されている。それらの方式の精度を比較する基盤として TrakMark ベンチマークテストがあり、TrakMark WG によって国際標準化を目指して策定作業が進められている。TrakMark WG による議論の中で、AR/MR 向けカメラキャリブレーションの精度を評価する際には、3次元空間内でのカメラ位置姿勢の誤差よりも、シーン中の様々な位置に仮想物体を置いた場合の画像上での表示位置の誤差を評価することが重要であり、そのために、シーン中に仮想的に3次元点群を配置し、その点群を画像上へ投影した時の誤差を評価するというアイデアが出された。本稿では、そのアイデアを定式化し、代表的な AR/MR 向けカメラキャリブレーション方式である PTAM を例に精度評価を行う。空間内でのカメラ位置姿勢の誤差と、PTAM が利用した特徴点の数の情報も同時に評価することで、提案した3次元点群の投影誤差による評価方式の意義について考察する。

キーワード 拡張現実感, 複合現実感, カメラキャリブレーション, TRAKMARK, PTAM, 誤差評価

Accuracy Evaluation of Camera Calibration for AR/MR by Projection Error of 3D Points in the Scene

Masayuki HAYASHI[†] Itaru KITAHARA[‡] Yoshinari KAMEDA[‡] and Yuichi OHTA[‡]

[†] Graduate School of Systems and Information Engineering, University of Tsukuba

1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8573 Japan

E-mail: [†] mhayashi@image.iit.tsukuba.ac.jp, [‡] {kitahara, kameda, ohta}@iit.tsukuba.ac.jp

Abstract Geometric consistency is one of the most important issue in Augmented-Reality and Mixed-Reality. TrakMark working group has provided benchmark sets for accurate evaluation of camera calibration methods. They also provide an idea of accuracy evaluation where virtual 3D points which are arranged at certain distance from the camera are projected on image plane to measure the geometric inconsistency. In this paper, we formulate the idea and evaluate the accuracy of Parallel Tracking and Mapping (PTAM) as an example.

Keyword Augmented-Reality, Mixed-Reality, TRAKMARK, PTAM, Error evaluation

1. はじめに

拡張現実感 (Augmented Reality; AR) や複合現実感 (Mixed Reality; MR) において、幾何学的整合性の確保は重要な課題である。そのために、様々なカメラキャリブレーション方式が提案されてきた[1]。しかし、これらの方式の性能を評価・比較する手法は現在模索されている状況である。

現在、TrakMark ワーキンググループ (TrakMark WG) により、国際標準化に向けたベンチマークテストの策定作業が行なわれている[2]。これまでに、異なるシーンを対象とした8種類のパッケージが公開されている。各パッケージには、特定のシーンを撮影したいいくつかの映像シーケンスと、撮影カメラの内部パラメータ、

及び各フレームにおける外部パラメータの情報が含まれている。様々なカメラキャリブレーション方式をこのデータに対して実行することで、公平な比較が可能になると期待されている。しかし、キャリブレーション精度の評価方法自体は未だ確立されていない。

カメラキャリブレーション方式には、画像中の特徴点の3次元位置を求め、その再投影誤差を最小化するカメラ位置姿勢を推定する方式がしばしば用いられる。この方法では、最小化された再投影誤差がキャリブレーション精度を評価する指標である。しかし、この指標では、特徴点の3次元位置付近に仮想物体を置いた場合の投影誤差を知り見することはできても、それ以外の位置に置いた場合に生じる投影誤差を予想するこ

とは難しい。

誤差評価の別の方式として、カメラの3次元位置と姿勢の誤差を距離差と角度差で表現する方法が提案されている[3,4]。空間内でのカメラ位置姿勢が正確であれば、空間中のどの位置に仮想物体を置いた場合でも、正しく投影されることが保証される。しかし、推定したカメラ位置姿勢に誤差が含まれていた場合、その誤差が画面上での仮想物体の表示位置にどの程度の影響を与えるのかを予想するのは難しい。

そこで、画面上での表示位置の誤差を評価するために、シーン中の様々な位置に仮想的に3次元点群を配置し、その仮想点群を参照カメラ、推定カメラの各々の画像上に投影し、その投影点同士の画像上での距離を評価する方式が TrakMark WG の議論から生まれた。

本稿ではまず、この評価方式の定式化を行う。次に TrakMark データセットを用いて、代表的な AR/MR の向けカメラキャリブレーション方式である PTAM[1] を例として精度評価を行う。空間内でのカメラ位置姿勢の誤差と、PTAM が利用した特徴点の数も同時に評価することで、3次元点群の投影誤差による提案評価方式の意義について考察する。

2. 3次元点群の投影誤差による評価

AR/MR における幾何学的整合性とは、仮想物体を現実映像中に正しい位置・姿勢で重畳表示することである。これは、現実空間における視点（カメラ）の位置姿勢を正確に推定することで達成できるため、カメラキャリブレーションの精度を評価する方法として、3次元空間内でのカメラ位置姿勢の精度を評価する手法がこれまで用いられてきた[3,4]。

しかし、実際には、空間内でのカメラ位置姿勢よりも、実写映像中の仮想物体の表示位置がどの程度ずれるかが重要である。そこで、シーン中に仮想的に3次元点群を配置し、その点群を参照値と推定値のカメラ位置姿勢を用いて画像上に投影した場合の投影誤差を評価する方式を提案する。

この方式を、図 1 に示す透視投影モデルを用いて定式化する。前提として、カメラの位置姿勢は、世界座標系からカメラ座標系への 4×4 の変換行列 \mathbf{C} として表現し、参照値と推定値をそれぞれ \mathbf{C}_t , \mathbf{C}_e とする。カメラ行列 \mathbf{C}_t , \mathbf{C}_e で表現されるカメラをそれぞれ参照カメラ、推定カメラと呼ぶ。

まず図 1 上に示すように、参照カメラの正規化画像座標上に N 点の格子点 \mathbf{x}_i , $i = 1, \dots, N$ とする。参照カメラの投影中心から、その格子点 \mathbf{x}_i を通る直線と、投影面に平行かつ投影中心から距離 d 離れた仮想平面 \mathbf{L} との交点 \mathbf{P}_i をシーン中の仮想的な3次元点群として配置する。これはつまり、 \mathbf{L} の近傍に仮想物体を重畳させることを期待している。この3次元点群 \mathbf{P}_i を推定カメラ

に投影し、正規化画像座標 \mathbf{y}_i を得る。

\mathbf{C}_t と \mathbf{C}_e が異なる場合、投影点 \mathbf{y}_i は格子点 \mathbf{x}_i からずれる。カメラの内部パラメータとレンズ歪みモデルを用いて \mathbf{x}_i , \mathbf{y}_i を画像座標に変換したものを \mathbf{x}'_i , \mathbf{y}'_i とする。このずれの平均を評価値とする。それらの距離 e_i の平均を平均投影誤差 E とする。

$$e_i = \|\mathbf{x}'_i - \mathbf{y}'_i\|$$

$$E = \frac{1}{N} \sum_i e_i$$

以上の手順によって、参照カメラから距離 d の仮想平面 \mathbf{L} 上においた仮想物体の表示位置のずれを求めることができる。参照カメラから仮想平面 \mathbf{L} までの距離 d を数パターン用意し、それぞれで平均投影誤差 E を求めることで、空間中の様々な奥行きに仮想物体を置いた時の表示位置ずれを評価することができる。

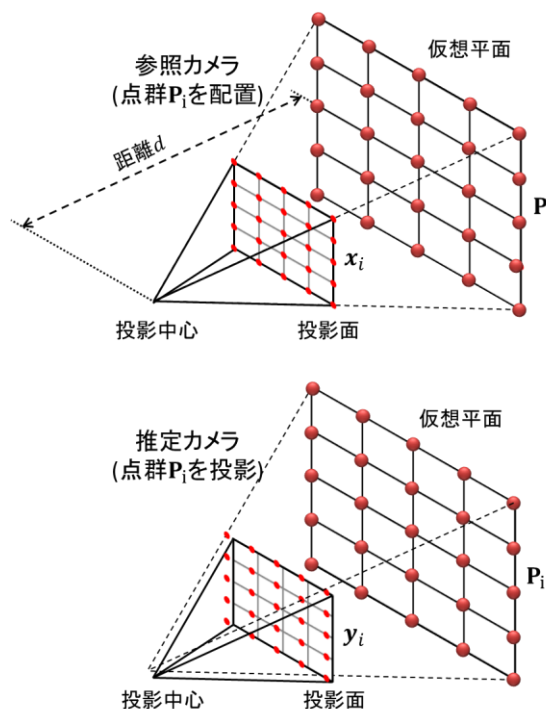


図 1 3次元点群の配置と推定カメラへの投影
Fig. 1 Projection of 3D points build from truth camera to estimated camera location

3. PTAM の評価

AR/MR 向けのマーカレスカメラキャリブレーション方式の代表的なアプローチとして、SLAM (Simultaneous Localization and Mapping) がある。特に、Klein ら[1]の提案した PTAM (Parallel Tracking and Mapping) は、未知の環境に対する特徴点の3次元マップをリアルタイムに構築し、その再投影誤差を最小化

するカメラ位置姿勢をリアルタイムに求めることができる。また、ソースコードも公開されている¹ため、AR/MR においてしばしば用いられるカメラキャリブレーション方式である[5]。

3.1. 評価手順

本稿では、以下の手順を用いて PTAM を提案方式で評価してみる。これは[4]で用いられている手順とほぼ同じである。

1. 内部パラメータの設定

PTAM が用いる内部パラメータは、正規化焦点距離 (f_x, f_y) 、正規化画像中心 (u_0, v_0) と Field of View モデル[5]の歪みパラメータ ω である。これらの情報は TrakMark データセットに含まれている。

2. ステレオ初期化に用いるフレームの設定

PTAM は初期の 3 次元特徴点マップを構築するためにステレオ視を用いる。そのために用いる 2 フレームを、シーケンス毎に決める必要がある。今回は先頭数十フレームの中から、初期化に成功しやすい 2 フレームを選択した。

3. PTAM の実行

映像シーケンスの各フレームにはタイムスタンプが付与されており、撮影時のフレームレートで入力することも可能であるが、本稿では PTAM の処理時間を確保するために、10 fps で入力を行った。

4. 世界座標系の変換

3.1.節の手順を用いて世界座標系の変換を行う。画像から推定されるカメラ位置はスケールが不定であるため、スケール付きユークリッド変換を行う。スケール s の計算方法は[4]と異なる。

5. 3次元点群の投影誤差による評価の実施

仮想平面を 4 パターン用意し、各フレームにおける平均投影誤差を求める。

3.2. 世界座標系の変換

PTAM によって推定されるカメラ行列は、ステレオ初期化時に設定された世界座標系を基準としている。そのため、参照値が基準としている実空間に固定された世界座標系とは、原点、姿勢、及びスケールが異なる。そこで、スケール付きユークリッド変換を行うことで参照値と比較可能なカメラ行列を得る。

ステレオ初期化直後のカメラ位置姿勢は正しく推定されていると仮定する。その時のカメラ行列の推定値 C_{e0} から参照値 C_{t0} への変換を T とする。

$$T = C_{t0}^{-1} \cdot sI \cdot C_{e0}$$

ここで、 I は 4×4 の単位行列である。[4]ではカメラが初期位置とそこから最も離れた位置の距離からスケール s を求めている。しかし、カメラ位置の推定誤差による影響を受けやすいため、本稿では PTAM が仮定するステレオ視のベースライン長 d_s と、それに対応する参照値 d_r の比を用いる。

$$s = d_r/d_s$$

最終的に、第 i フレームにおけるカメラ行列 C_{ei} から変換行列 T を用いて、参照値 C_{ti} と比較可能なカメラ行列 C'_{ei} を得る。

$$C'_{ei} = C_{ei} \cdot T^{-1}$$

4. 実験結果

3 節の手順を用いて PTAM のキャリブレーション精度評価を行った結果を示す。今回はカメラの投影面上に 17×17 点の格子点を用いて配置した 3 次元点群を用いた ($N=289$)。

今回は実行結果として、空間内でのカメラ位置、姿勢 (オイラー角) の参照値 (TX, TY, TZ) と推定値 (X, Y, Z)、3 次元点群の投影誤差に加えて、PTAM が構築した 3 次元マップの特徴点数 (nmp) とキーフレーム数 (nkf)、画像中の特徴点との 2D-3D 対応付けを試みた特徴点数 (apt)、実際に対応付けができた特徴点数 (fnd) を同時に観察して評価する。

各パッケージの結果と考察は以下のとおりである。

4.1. Conference Venue Package 01

このパッケージに含まれる映像シーケンスは全て、CG による屋内シーンであり、カメラの動きが滑らかである。

カメラが移動と回転を行う Translation + Panning + Tilting シーケンスのサンプル画像と、結果を図 2 に示す。解像度は 640×480 画素である。フレーム間の姿勢変動が少なく、シーケンス全体に渡って比較的高精度なカメラキャリブレーションが行えていることがカメラ位置、姿勢のグラフよりわかる。

3 次元点群の平均投影誤差 E は $d = 1, 2, 5, 10$ [m] とした時の 3 次元点群を用いて評価した。結果を見ると、最大でも 9 画素程度に抑えられていることがわかる。また、点群までの距離が 1 [m] の時に誤差が大きく、それより遠い点群では誤差が小さいことがわかる。これは、PTAM が用いた特徴点のほとんどが壁上にあるため、壁より手前の空間は投影誤差が生じやすかったためと考えられる。

さらに、キーフレーム数 (nkf) とマップ点数 (nmp) は単調増加しているが、apt や fnd が減少しているフレーム (300 フレーム付近, 1000 フレーム以降) において、平均投影誤差 E が増加していることがわかる。

4.2. NAIST Campus Package 02

このパッケージに含まれる映像シーケンスは、屋外

¹ Parallel Tracking and Mapping for Small AR Work spaces - Source Code. <http://www.robots.ox.ac.uk/~gk/PTAM/>

歩行者が持つカメラからの映像で、解像度は 1280x720 画素である。

Sequence 00 のサンプル画像と結果を図 3 に示す。このシーケンスでは、特徴点は十分な数が抽出された。シーケンス前半では再投影誤差が小さい特徴点が多くあり、そのような特徴点は主に遠方の建物上にあった。そのため、建物上に仮想物体を設置する場合は、画像内での幾何的整合性が確保されると考えられる。しかし、カメラ位置の推定誤差を見ると、姿勢や X 軸、Y 軸方向の位置の誤差はあまり大きくないものの、奥行き方向 (Z 軸) の誤差が大きくなっていることがわかる。

3次元点群の平均投影誤差は、撮影シーンの広さに合わせて $d = 2, 5, 10, 50$ [m]とした時の3次元点群を用いて評価した。その結果、投影誤差は非常に大きく、カメラ近くに用意した仮想点群ほど誤差が大きくなった。これは、ほとんどの特徴点が遠くの建物上であったため、それより手前の空間で投影誤差が生じやすいためと考えられる。

これは、Z 軸方向のカメラ位置の推定誤差が大きく、その分だけ仮想点群が推定カメラから遠くなり、投影点が画面中央付近に集中する。どの距離の仮想点群においても、1800 フレーム付近までは、投影点が画面中央に集中した時の投影誤差に近づいている。その後、カメラの向きが変わり姿勢の推定誤差が大きくなったため、投影点が画面外に大きく外れ、投影誤差が増えている。

nkf や nmp を見ると、対応付けを試みた点数 apt は PTAM の参照実装の上限である 1000 点程あるにも関わらず、実際に対応付けができた数 fnd は少いことがわかる。

4.3. Film Studio Package 02

このパッケージに含まれる映像シーケンスは、屋内のフィルムスタジオで撮影されたハイビジョンカメラ映像で、解像度は 1920x1080 画素である。

カメラが移動と回転を行う Sequence 04 のサンプル画像と結果を図 4 に示す。カメラの位置、姿勢とも推定誤差が生じているが、壁や床の特徴点には再投影誤差が小さい点も十分にあり、カメラトラッキングは破綻していなかった。

3次元点群の平均投影誤差 E は $d = 1, 2, 5, 10$ [m]とした時の3次元点群を用いて評価した。その結果、1[m]の点群で最大 900 画素程度 (300 フレーム付近) と大きく、仮想物体をカメラ近くに置くとその表示位置が画面外になる可能性が高いといえる。一方、 $d = 5$ [m]の仮想点群の平均再投影誤差 E は 200 画素以下に抑えられていた。PTAM はカメラ位置姿勢を推定する際、2D-3D 対応の取れた特徴点の再投影誤差を最小化する

ため、その特徴点付近に仮想点群がある場合は投影誤差が小さくなると考えられる。今後、推定した特徴点の3次元位置の分布を観察して確かめる必要がある。

この映像シーケンスでは、マップに含まれる特徴点の数は十分にあり、対応付けを試みた点数 apt は約 1000 点あるが、実際に対応付けができた数 fnd は半分以下になる場合が多かった。

5. まとめ

TrakMark WG によってアイデアが出された、シーン中の仮想的な3次元点群の投影誤差によるカメラキャリブレーションの精度評価方式を定式化した。この方式を用いて、代表的な AR/MR 向けカメラキャリブレーション方式である PTAM を例に評価を行った。その結果、提案評価方式は、空間内での仮想物体の位置と、画像中での表示位置の誤差との関係を示すことができた。

提案した評価方式は、画像内での幾何学的整合性のよさを直接知ることができ、複数の AR/MR 向けカメラキャリブレーション手法の精度を比較しやすい評価指標であると考えられる。

なお、本研究で用いたプログラムは、USB boot 可能な Linux による Casper Cartridge[6]を用いて公開予定である²。

謝 辞

本研究の基本的なアイデアと、その評価指標としての意義に関する数々のアドバイスを下さった TrakMark Working Group の委員の方々に、感謝いたします。

本研究の一部は、日本学術振興会基盤研究(B) (課題番号 23300064) および平成 23 年度科学研究費補助金 (特別研究員奨励費: 課題番号 23・310) の支援を受けて行われた。

文 献

- [1] G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," *6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1–10, 2007.
- [2] H. Tamura and H. Kato, "Proposal of international voluntary activities on establishing benchmark test schemes for AR/MR geometric registration and tracking methods," *8th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 233–236, 2009.

² Casper Cartridge Project: <http://www.kameda-lab.org/casper/index-j.html>

- [3] A. Petit and G. Caron, "Evaluation of Model based Tracking with TrakMark Dataset," *The 2nd International Workshop on AR/MR Registration, Tracking and Benchmarking (TrakMark2011)*, 5 pages, 2011.
- [4] M. Hayashi, "A Study of Camera Tracking Evaluation on TrakMark Data-Set," *The 2nd International Workshop on AR/MR Registration, Tracking and Benchmarking (TrakMark2011)*, 5 pages, 2011.
- [5] R. a. Newcombe and A. J. Davison, "Live dense reconstruction with a single moving camera," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1498–1505, 2010.
- [6] 林, 北原, 亀田, and 大田, "AR/MR デモンストレーションの再現性を保証する ソフトウェア環境の構築 -Casper Cartridge プロジェクター," *電子情報通信学会技術研究報告. MVE, マルチメディア・仮想環境基礎*, vol. 111, no. 235, pp. 103–108, 2011.

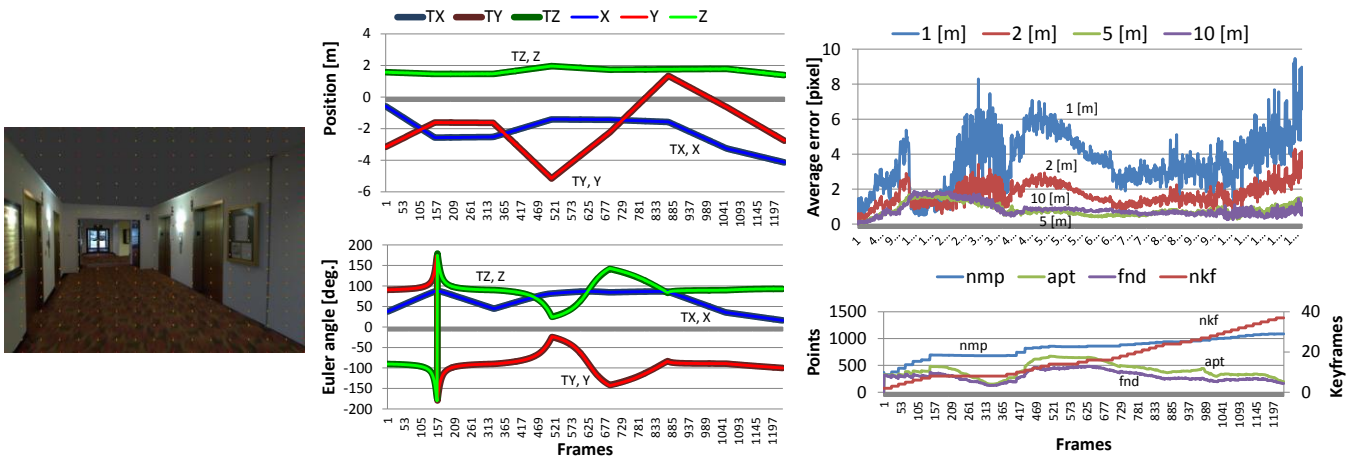


図 2 Conference Venue Package 01 の実行結果. サンプル画像 (左), 空間内でのカメラ位置姿勢 (中央), 3次元点の投影誤差 (右上), 特徴点数 (右下).

Fig. 2 Result of Conference Venue Package 01. Sample image (left), camera position and orientation in the scene (center), projection error of 3D points (upper-right), number of feature points (lower-right).

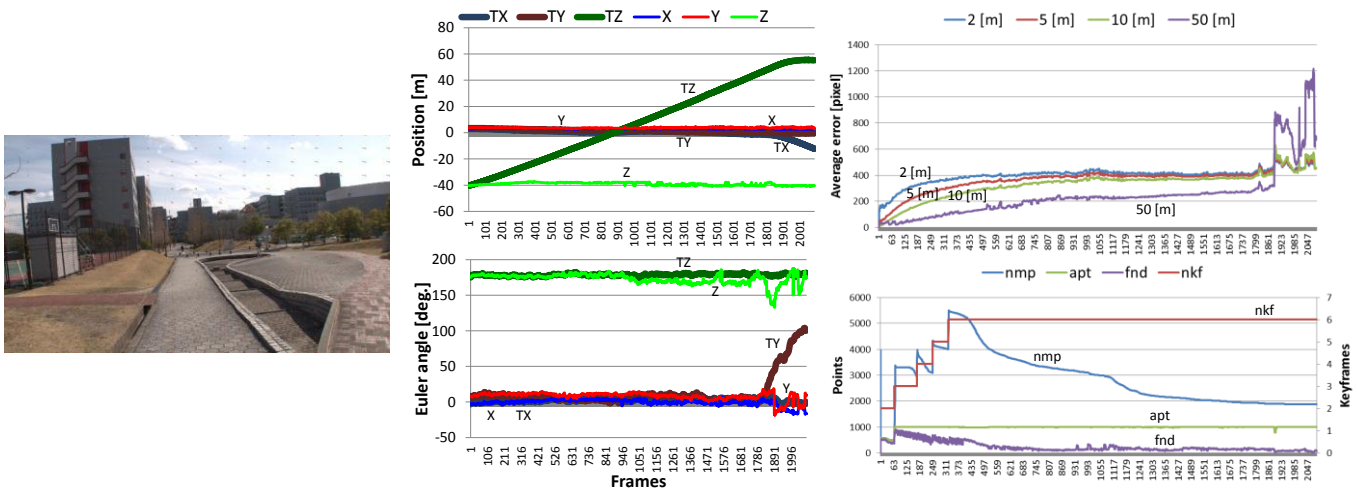


図 3 NAIST Campus Package 02 の実行結果. サンプル画像 (左), 空間内でのカメラ位置姿勢 (中央), 3次元点の投影誤差 (右上), 特徴点数 (右下).

Fig. 3 Result of NAIST Campus Package 02. Sample image (left), camera position and orientation in the scene (center), projection error of 3D points (upper-right), number of feature points (lower-right).

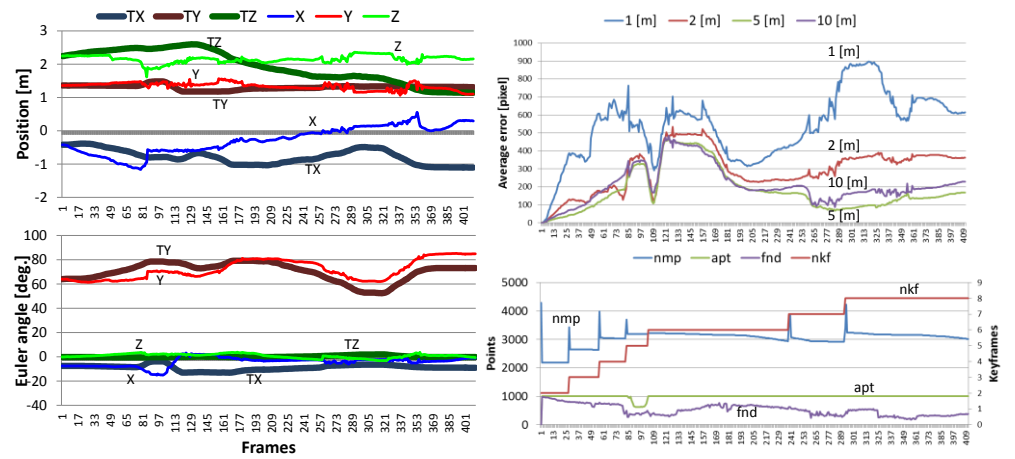


図 4 Film Studio Package 02 の実行結果．サンプル画像（左），空間内でのカメラ位置姿勢（中央），3次元点の投影誤差（右上），特徴点数（右下）．

Fig. 4 Result of Film Studio Package 02. Sample image (left), camera position and orientation in the scene (center), projection error of 3D points (upper-right), number of feature points (lower-right).